

KLASTERISASI PROSES SELEKSI PEMAIN MENGGUNAKAN ALGORITMA K-MEANS

(Study Kasus : Tim Hockey Kabupaten Kendal)

Alith Fajar Muhammad

Jurusan Teknik Informatika FIK UDINUS, Jl. Nakula No. 5-11 Semarang-50131

alithfm@gmail.com

Abstrak - Klasterisasi pemain hockey dilakukan untuk mengelompokkan pemain kedalam cluster tertentu yang memiliki kemiripan data menggunakan algoritma K-means. Data yang diambil meliputi pukulan hit, push, tapping serta data *Multi level running speed* dan Sprint 50 meter. Metode evaluasi cluster menggunakan DBI dan purity untuk mengukur seberapa bagus cluster yang dihasilkan. Dari hasil penelitian yang menggunakan 100 data pemain (59 Putra 41 Putri) menghasilkan 3 cluster dengan evaluasi nilai sebagai berikut : Untuk nilai DBI pada Cluster Putra adalah 0.95934206 dan Cluster Putri adalah 0.976979445. Sedangkan Nilai Purity untuk Cluster Putra adalah 0.879 dan Cluster Putri adalah 0.608. Ini membuktikan bahwa pengelompokan dengan k-means menghasilkan cluster yang masih belum cukup maksimal.

Kata kunci : Hockey, K-means, Cluster , Matlab, *davies-bouldin index* (DBI), Purity

I. PENDAHULUAN

Datamining berisi pencarian trend atau pola yang diinginkan dalam database yang besar untuk membantu pengambilan keputusan diwaktu yang akan datang. Harapannya, perangkat *datamining* mampu mengenali pola-pola ini dalam data dengan masukan yang minimal. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya [1].

Clustering (Klasterisasi) adalah proses mengelompokkan atau penggolongan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas / *cluster*. *Clustering* dalam *datamining* berguna untuk menemukan pola distribusi di dalam sebuah data set yang berguna untuk proses analisa data. Kesamaan objek biasanya

diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangkan objek-objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi [2].

Hockey adalah olahraga permainan yang dilakukan oleh pria dan wanita dengan menggunakan alat pemukul (*stick*) dan bola. [3]. Dari uraian diatas olahraga *hockey* yang populer di jawa tengah ada di beberapa kota/kabupaten. Salah satunya adalah tim *hockey* kabupaten kendal. Tim *hockey* Kabupaten Kendal lahir pada tahun 2005. Seiring dengan pesatnya olahraga *hockey* di kabupaten kendal khususnya di kecamatan boja dan sekitarnya sekarang jumlah anggotanya lebih dari 50 orang, terdiri dari pria dan wanita.

Pada proses seleksinya Tim Kepelatihan tidak selalu menyeleksi pemain berdasarkan kriteria yang ditetapkan, sehingga terkadang proses seleksi tidak berjalan sesuai kemampuan yang diukur dengan tepat untuk setiap pemain.

Oleh karena itu diadakan penelitian yang bertujuan untuk menggali potensi data personal di setiap pemain yang akan di kelompokkan ke dalam *cluster* menggunakan algoritma K-Means.

II. TEORI PENUNJANG

2.1 Metode Clustering

Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan (kemiripan). *Clustering* adalah proses mengelompokkan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/*cluster*.

Tujuannya menemukan *cluster* yang berkualitas dalam waktu yang layak. *Clustering* dalam *datamining* berguna untuk menemukan pola distribusi di dalam sebuah data set yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangkan objek-objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi [7].

Beberapa manfaat clustering adalah sebagai berikut :

- a. Identifikasi Obyek (*Recognition*)
Dalam bidang *image processing*, *Computer Vision* atau robot vision.
- b. *Decission Support System* (Segementasi Pasar)

2.2 Algoritma K-Means

Algoritma *k-means* merupakan algoritma yang digunakan untuk pengelompokan iteratif, algoritma ini melakukan partisi set data ke dalam sejumlah K cluster yang sudah ditetapkan diawal. Partisi set data tersebut dilakukan untuk mengetahui karakteristik dari setiap *cluster*, sehingga *cluster* yang memiliki karakteristik sama dikelompokkan

kedalam satu *cluster* dan yang memiliki karakteristik berbeda dikelompokkan kedalam *cluster* lain. Berikut langkah-langkah perhitungan dalam *k-means*, antara lain [1] [3] [9] [7] [10] [11]:

1. Tentukan jumlah *cluster* dan ambang batas perubahan fungsi objektif
2. Menentukan *centroid* awal yang digunakan
3. Menghitung jarak setiap data ke masing-masing *centroid* menggunakan jarak *euclidean* untuk mendapatkan jarak terdekat data dengan *centroid*nya
4. Menentukan *centroid* baru dengan menghitung nilai rata-rata dari data yang ada pada *centroid* yang sama
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu perubahan fungsi objektif sudah dibawah ambang batas yang diinginkan, atau tidak ada data yang berpindah *cluster*, atau perubahan posisi *centroid* sudah dibawah ambang batas yang sudah ditentukan.

Setelah perhitungan jarak dari setiap data terhadap *centroid* dihitung, kemudian dipilih jarak yang paling kecil atau yang mendekati nilai 0 sebagai *cluster* yang akan diikuti sebagai relokasi data pada *cluster* di sebuah iterasi. Relokasi sebuah data dalam cluster yang diikuti dapat dinyatakan dengan nilai keanggotaan *a* yang bernilai 0 atau 1. Nilai 0 jika tidak menjadi anggota sebuah cluster dan 1 jika menjadi anggota sebuah cluster. *K-means* mengelompokkan data secara tegas hanya pada satu cluster, maka nilai *a* sebuah data pada semua *cluster*, hanya satu yang bernilai 1. Perhitungan jarak antara data dan *centroid* dapat dilakukan dengan menggunakan persamaan *euclidean distance*, persamaannya sebagai berikut [1] [7] [10] [11] :

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.1)$$

Keterangan :

d_{ij} = Jarak objek antara objek i dan j

P = Dimensi data

x_{ik} = obyek i pada dimensi k

x_{jk} = obyek j pada dimensi k

Dan dibawah ini persamaan untuk mencari nilai fungsi objektif setiap data [1] [7]:

$$\sum_{li=1}^{NK} \{x_{li} - c_{li}\}^2 \quad (2.2)$$

NK adalah jumlah data yang tergabung dalam sebuah cluster

2.3 Evaluasi Cluster

Metode yang digunakan dalam menentukan evaluasi cluster ini menggunakan *davies-bouldin index (DBI)* dan *purity*.

1. Davies-Bouldin Index

Davies-bouldin index merupakan salah satu metode evaluasi internal yang mengukur evaluasi cluster pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. Dalam suatu pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap centroid dari cluster yang diikuti. Sedangkan separasi didasarkan pada jarak antar centroid dari clusternya.

Sum of square within cluster (SSW) merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah cluster ke-i yang dirumuskan sebagai berikut :

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (2.3)$$

Dari persamaan tersebut, m_i merupakan jumlah data dalam cluster ke-i, c_i adalah centroid cluster ke-i, dan $d()$ merupakan jarak setiap data kecentroid yang dihitung menggunakan jarak *euclidean*.

Sum of square between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar cluster yang dihitung menggunakan persamaan :

$$SSB_{i,j} = d(c_i, c_j) \quad (2.4)$$

Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara cluster ke-i dan cluster ke-j. Cluster yang baik adalah cluster yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Nilai rasio dihitung menggunakan persamaan sebagai berikut :

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (2.5)$$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *davies-bouldin index (DBI)* dari persamaan berikut :

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \quad (2.6)$$

Dari persamaan tersebut, k merupakan jumlah cluster yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik cluster yang diperoleh dari pengelompokan *K-means* yang digunakan [7].

2. Purity

Purity digunakan untuk menghitung kemurnian dari suatu cluster yang direpresentasikan sebagai anggota

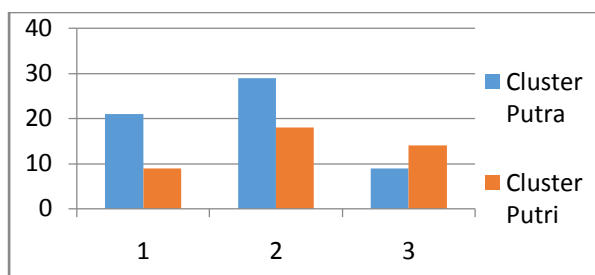
cluster yang paling banyak sesuai (cocok) disuatu kelas. Nilai *purity* yang semakin mendekati 1 menandakan semakin baik *cluster* yang diperoleh. Untuk menghitung nilai *purity* setiap *cluster* dapat menggunakan rumus berikut [13]:

$$Purity(j) = \frac{1}{n_j} \max_i(n_{ij}) \quad (2.7)$$

Sementara untuk menghitung *purity* keseluruhan jumlah K *cluster*, digunakan persamaan sebagai berikut:

$$Purity = \sum_{j=1}^k \frac{n_j}{n} Purity(j) \quad (2.8)$$

III. HASIL & IMPLEMENTASI



Hasil *cluster* untuk putra dan putri menunjukkan bahwa *Cluster 1* terdapat 21 putra dan 9 putri, *Cluster 2* terdapat 29 putra dan 18 putri sedangkan untuk *Cluster 3* terdapat 9 putra dan 14 putri, ini menunjukkan bahwa *cluster* yang lebih dominan memiliki kemiripan data adalah cluster ke 2.

DBI PUTRA 0.95934206	DBI PUTRI 0.976979445
PURITY PUTRA 0.879	PURITY PUTRI 0.608

Dari perhitungan tersebut telah diperoleh nilai *DBI* untuk Cluster Putra adalah 0.95934206 dan untuk Cluster putri adalah 0.976979445. Semakin kecil nilai

DBI atau semakin mendekati nilai 0 menunjukkan seberapa baik cluster yang diperoleh, sehingga nilai *DBI* yang telah diperoleh tersebut menunjukkan cluster dihitung dalam penelitian ini masih belum cukup bagus.

Sedangkan untuk Evaluasi *cluster* menggunakan *Purity* adalah sebagai berikut :

Putra					
Cluster	Jumlah	Bagus	Sedang	Kurang	Purity
C1	21	0	4	17	0.809
C2	29	3	26	0	0.896
C3	9	9	0	0	1
Total	59				
Purity Putra					0.879

Putri					
Cluster	Jumlah	Bagus	Sedang	Kurang	Purity
C1	9	1	6	2	0.666
C2	18	13	5	0	0.722
C3	14	2	6	6	0.428
Total	41				
Purity Putri					0.608

Dari perhitungan tersebut telah diperoleh nilai *purity* untuk Cluster Putra adalah 0.879 dan Cluster Putri adalah 0.608 . Hasil yang diperoleh ini menunjukkan nilai *purity* dari setiap data *cluster* yang dihasilkan cukup bagus untuk cluster putra karena mendekati nilai 1 dan belum cukup bagus untuk cluster putri karena belum mendekati nilai 1.

IV. PENUTUP

Berdasarkan penelitian yang dilakukan dalam klasterisasi proses seleksi pemain tim Hockey Kendal menggunakan algoritma K-means , menghasilkan Tiga Cluster yang telah tersusun untuk 100 Pemain (59 Putra dan 41 Putri) serta nilai evaluasi *DBI* serta *Purity* dari cluster tersebut. Sedangkan hasil dari penelitian yang diperoleh menunjukkan nilai *DBI* dan *purity* yang masih cukup rendah untuk dapat dikatakan mendekati nilai 0 atau 1. Sehingga dapat dikatakan proses Cluster yang dihasilkan belum cukup bagus untuk dikatakan sebagai cluster yang baik. Hasil ini dapat dipengaruhi berbagai hal diantaranya adalah proses pengukuran data

(tolok ukur) yang mengakibatkan nilai diantara pemain hampir sama.

DAFTAR PUSTAKA

- [1] T. Rismawan dan S. Kusumadewi, "Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index & Ukuran Kerangka," *SNATI*, pp. 43-48, Juni 2008.
- [2] H. A. Fajar, *DATA MINING*, vol. 2, Yogyakarta: ANDI, 2013.
- [3] A. Elizabeth dan M. Sue, *Field Hockey Step To Succes 2nd*, vol. 15, New Zealand: Human Kinetics, 2008, pp. 73-82.
- [4] Ediyanto, M. N. Mara dan N. Satyahadewi, "Pengklasifikasian Karakteristik dengan Metode K-Means Cluster Analysis," *Buletin Ilmiah, Mat Stat, dan Terapannya (Bimaster)*, vol. 02, pp. 133-136, 2013.
- [5] P. D. S. Sani dan M. S. Dedy ST, *Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data*, Jogjakarta: ANDI OFFSET, 2010.
- [6] T. Primadi, *Hockey dan Kreativitas dalam Olahraga*, Bandung: ITB, 1985.
- [7] S. C. M. CitraDevi dan G. Geetharamani, "An Analysis on The Performance of K-Means Clustering Algorithm For Cardiotogram Data Cluster," *International Journal on Computational Sciences & Applications*, vol. 2, no. 5, pp. 11-20, Oktober 2012.
- [8] Widiarina dan W. R. S., "Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam Pemetaan Nasabah Potensial," *Journal Of Intelligent System*, vol. 1, no. 1, pp. 32-35, Februari 2015.
- [9] R. Handoyo, R. R. M. dan S. M. Nasution, "Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means pada Pengelompokan Dokumen," *JSM STMK Mikrosil*, vol. 15, no. 2, pp. 73-82, Oktober 2015.
- [10] T. Khotimah, "Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means," *Jurnal SIMETRIS*, vol. 5, no. 1, pp. 83-88, April 2014.
- [11] J. Ong, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University," *Jurnal Ilmiah Teknik Industri*, vol. 12, no. 1, pp. 10-20, Juni 2013.
- [12] P. Eko, *DATA MINING. Mengolah Data Menjadi Informasi Menggunakan Matlab*, Yogyakarta: ANDI, 2014.
- [13] K. Prilianti dan H. Wijaya, "Aplikasi Text Mining Untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering," *Jurnal Cybermatika*, vol. 2, no. 1, Juni 2014.
- [14] T. E. Purwoastuti dan W. S. Elishabet, *Metodologi Penelitian*, Yogyakarta: PUSTAKABARUPRESS, 2014.